The Nucleus **47**, No. 3 (2010) 211-217

The Nucleus
A Quarterly Scientific Journal of Pakistan Atomic Energy Commission
NCLEAM, ISSN 0029-5698

COMPARISON OF AUTOREGRESSIVE (AR) STRATEGY WITH THAT OF REGRESSION APPROACH FOR DETERMINING OZONE LAYER DEPLETION AS A PHYSICAL PROCESS

*M.A.K. YOUSUF ZAI, R.K. ANSARI, J. QUAMAR, J. IQBAL and M.A. HUSSAIN

Department of Applied Physics and Institute of Space and Planetary Astrophysics, University of Karachi, Karachi-75270, Pakistan

(Received January 22, 2008 and accepted in revised form July 27, 2010)

This communication presents the development of a comprehensive characterization of ozone layer depletion (OLD) phenomenon as a physical process in the form of mathematical models that comprise the usual regression, multiple or polynomial regression and stochastic strategy. The relevance of these models has been illuminated using predicted values of different parameters under a changing environment. The information obtained from such analysis can be employed to alter the possible factors and variables to achieve optimum performance. This kind of analysis initiates a study towards formulating the phenomenon of OLD as a physical process with special reference to the stratospheric region of Pakistan. The data presented here establishes that the Auto regressive (AR) nature of modeling OLD as a physical process is an appropriate scenario rather than using usual regression. The data reported in literature suggest quantitatively the OLD is occurring in our region. For this purpose we have modeled this phenomenon using the data recorded at the Geophysical Centre Quetta during the period 1960-1999. The predictions made by this analysis are useful for public, private and other relevant organizations.

Keywords: Autoregressive process, Ozone layer depletion, Multiple regression, Simple regression, Stochastic process, Atmospheric modeling

1. Introduction

The evolution of science has increased the understanding of various aspects of the environment and consequently the predictability of many naturally occurring events. It is surprising to note that in general, quantitative studies of the well-known problem of ozone layer depletion (OLD) are not available yet. This is particularly true for Pakistan. However, in the past few decades, there has emerged a widespread awareness of the importance of ozone from a new perspective that human activities may themselves be responsible for polluting the earth's stratosphere and consequently disturbing the ozone (O_3) layer balance [1-3]. Due to serious reduction of O_3 content in the atmosphere, life forms on earth are exposed to a new hazard in the form of an increase in harmful UV radiation at sea level [4,5,8]. This situation calls for an assessment, monitoring and prevention of the incidence of decrease in the O_3 concentration. This paper illustrates mathematical modeling of OLD phenomenon. In this study we have examined the quantitative measures and relationships for a specific portion of numeric data obtained via various observational programs on the global ozone network monitored by WMO stations [6,7,9].

This communication presents the development of a comprehensive characterization of ozone layer depletion (OLD) phenomenon as a physical process in the form of mathematical models that comprise the usual regression, multiple regression and stochastic strategy. The relevance of these models has been illuminated using predicted values of different parameters under a changing environment [10]. The information obtained from such analysis can be employed to alter the possible factors and variables in the process of OLD to achieve an optimum performance. This kind of analysis initiates a study towards formulating the phenomenon of OLD as a physical process with special reference to the stratospheric region of Pakistan.

2. Modeling Strategies

The results presented in this publication establish that Auto regressive (AR) nature of

* Corresponding author : ayubzai@yahoo.com

modeling OLD as a physical process is an appropriate scenario rather than using regression. The reported data suggest quantitatively the phenomenon of OLD is occurring. For this purpose we have modeled this phenomenon using the data recorded at the Geophysical Center Quetta during the period 1960-1999.

(a) Regression analysis describes the estimation of the unknown value of one variable from the known value of the other variable. In regression analysis there are two types of variables. First type tells us that the variable whose value is influenced or is to be predicted is called the dependent variable and the variable which influences the values or is used for the prediction is called the independent variable. Prediction is based on the average relationship derived statistically by regression analysis. In particular, regression is the measure of the average relationship between two or more variables in terms of the original units of data.

Figure1. is a plot of O_3 layer width *H* in Dobson units against time *t* in months. The plot appears to be a straight line having negative slope. This relationship may represented by

$$H = a + bt \qquad (1)$$

Where H is the O_3 depth and t is the time. In terms of a pattern and an error equation it can be represented as

$$H = (\text{pattern}) + (\text{error}) \equiv \hat{y} + (\text{error}) \qquad (2)$$

The observed value of the ozone will differ from the predicted value by error term (ε) that is unknown. Albeit we can rather state the exact value of the ozone concentration, we can make a probabilistic statement about it using the predicted value and fact that ε_t ~ Normally Independently Distributed $(0, \sigma^2)$.

For our sample data consisting of 12 observations we have

n_l = 12,

a = 329.25 (parameter value), intercept

b = −2. 51 (parameter value), slope of the line

$\hat{y} = 329.25 - 2.51\,x$

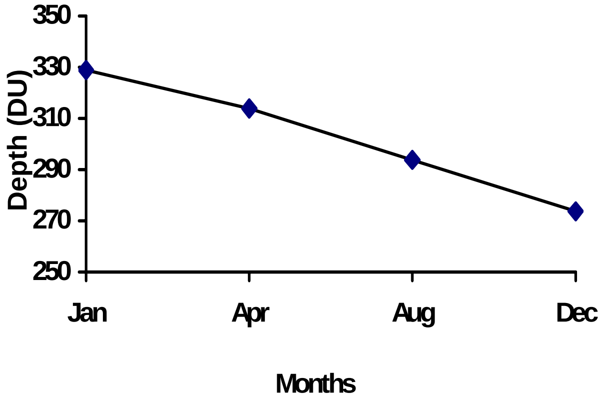

Figure 1. Temporal trend by least square method for the year 1982

Table 1. Trend analysis by least square method for Ozone layer for the year 1982.

Months	T	H	t^2	t H
January	−11	329	121	−3619
February	−9	322	81	−2898
March	−7	320	49	−2240
April	−5	316	25	−1580
May	−3	315	9	−945
June	−1	305	1	−305
July	1	299	1	299
August	3	291	9	873
September	5	294	25	1470
October	7	276	49	1932
November	9	270	81	2430
December	11	279	121	3069

Some tentative predictions for Pakistan have been listed using the same concept as in the year 1996 the ozone level was 292.82 DU and for the year 1997 it was 292.40 DU. It is important to realize that in many atmospheric data sets assumptions such as randomness may well be violated simply because of structural nature of the information. The statistical procedures may on occasion be modified to account for some of the violations caused by the atmosphere. In view of the above discussion our results recorded above establish that the O_3 layer depletion is a threat to mankind and it is also affecting Pakistan's environment. The above model seems to have practical implication and helpful in making predictions for various national and international organizations. Underlying these measures are assertions about the probability distributions and assumptions of randomness, independence and constant variance. These and some of their variants will be tackled in future.

(b) A stochasti**c** or time series analysis describes the arrangement of the data in accordance with the time of occurrence. If the values of a phenomenon are observed at different periods of time or events, the values so attained will exhibit appreciable or reasonable variations. It helps in analyzing the phenomenon in terms of the effect of various technological, economic and other factors on its behaviour over time. This also helps in forecasting a most likely value of a variable in near future [11-12]. The desirable goal of this section is to develop a comprehensive characterization of the ozone layer depletion phenomenon as a physical process in the form of mathematical models.

The goal of this section is to model OLD phenomenon in such a way that a close resemblance can be established between the consequent predictions and the associated empirical data. For this purpose we will construct autoregressive models of various orders so that an appropriate model could be selected to determine that OLD is a physical process. Figure 2 shows temporal stationary but random variations of ozone depths that were recorded at the Geophysical Centre, Quetta, Pakistan where Dobson Spectrophotometer has been installed under auspices of World Meteorological Organization (WMO).

(c) Construction of time series model to predict OLD for Pakistan's atmospheric region.

It has been known that an important aspect of scientific study is crystallized by the idea of a model. A model can define the real situation of a system. We are having the data which consists of continuous variables. An observed time series can be thought of a particular realization of Stochastic. Time series is defined by a record of the values of any fluctuating quantity measured at different points of time. We may, for example, have a record of the depths of O_3 layer over a period of 40 years (1960-1999) which is being used in this study.

(d) Criteria for model selection of OLD in Pakistan's atmospheric region.

The common feature of all the records which fall within the domain of ' time series analysis' is that they are influenced, at least in part, by the random variations, as the O_3 layer depth is affected by certain atmospheric events. Thus if we intend to explain particular pattern of the fluctuations in the O_3 depths in the stratospheric region of Pakistan, then we need to construct a mathematical description of the OLD data obtained from a WMO installation of Dobson Spectrophotometer. Such a model will explain both the deterministic and random features of the OLD.

The construction of the model is one of the basic objectives of OLD analysis, if we can obtain an adequate model for our series, it may provide valuable insight into the physical mechanism generating the data, and it can be used to forecast, for example the future values of the series.

To understand the concept we can illustrate the case of autoregressive model which is frequently used time series model. Autoregressive model is the special case of multiple regression model in which some or all the explanatory variables are lagged values of X_t.

The autoregressive model takes the form

$$X_t = \alpha_o + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_S X_{t-S} + e_t \quad (3)$$

The value X_{t-S} is called the lagged value of Y at time (t-s). The order of autoregressive model is s.

where X_t is expressed as a linear combination of its two immediately preceding values. The value variables X_{t-1} and X_{t-2} are constructed easily by moving the values down the available data. Auto-corrrelation between X_t and X_{t-1} and between X_t and X_{t-2} can be computed.

$$X_t = \phi X_{t-1} + \alpha_o \quad (4)$$

Figure 2 is a monthly original time plot of ozone depth data, and Figure 3 is a scatter plot of the data. From this graph it appears that knowing the value of ozone content in the period (t-1) is useful in predicting the value of the total value of ozone concentration in period t. It seems that X_t can be explained as function of X_{t-1}. Now estimate the regression coefficient of order one.

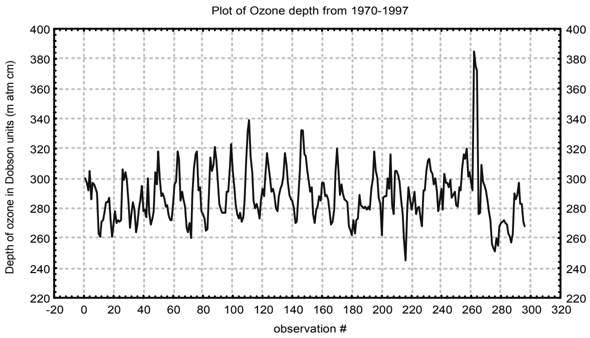

Figure 2. Variations of ozone depths versus time, showing that the depletion far exceeds the restoration of O_3 in the earth's stratosphere.

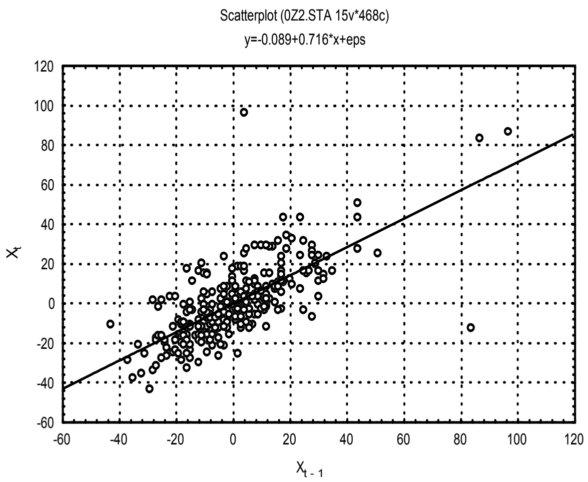

Figure 3. Ozone depth data: X_{t-1} versus X_t.

$x_t = \phi_1 X_{t-1} + \alpha_o$

$\hat{x}_t = 81.772 + 0.716\, x_{t-1}$

t statistics: (6.93) (17.46) R^2 = 51 percent

where α_o = 81.772, ϕ_1= 0.716, and it shows that $\phi_1 < 1$.

The t statistics for ϕ_1 is 17.46 and the value of p = 0.0000 for ϕ_1. The t statistics in an autoregressive model does not exactly follow the t distribution because one of the basic assumptions of the classical linear regression model has been violated.

From the analysis of variance (ANOVA), the coefficient of determination (R^2) can be calculated as

$$R^2 = \frac{SS}{\text{Total SS}} = \frac{50825}{99649} = 0.51$$

This indicates that 51 % of the variation is explained by the regression model. The remaining 49 % of the variation is itself unexplained. It shows that a good forecast of the value of X is possible when the previous `value of X is known. Using equation (42) to forecast the ozone depth for 297th month by substituting x_{296} = 268 DU. From the equation , we obtain

$\hat{x}_{297} = 81.772 + 0.716\, x_{296}$

x_{296} = 268 DU, and obtain

$\hat{x}_{297}$ = 81.772 + 0.716 (268) = 273.66 DU

Ozone depth is measured in Dobson unit (1 DU = 10^{-3} cm of O_3 at STP of the atmosphere).[1]. Thus the forecast for the 297th of the period specified for this communication is 273.66 DU.

The periodogram illustrated in Figure. 4, is used to identify randomness in the data series. Also it helps in identifying seasonality in the given time series, and in recognising the predominance of positive or negative autocorrelation (for positive autocorrelation low-frequency amplitudes should dominate, and for negative autocorrelation, high frequencies should dominate).

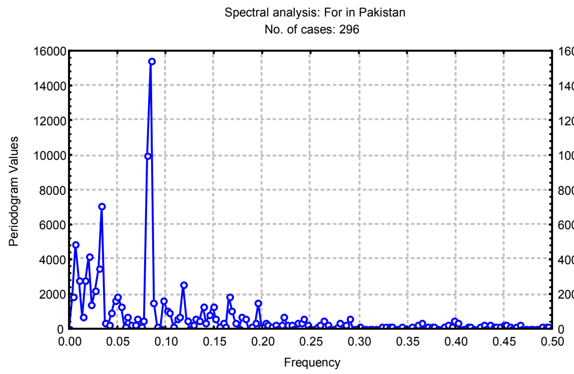

Figure 4. Periodogram as shown is used to identify randomness in the OLD data.

i. Autoregressive processes may be revealed—pattern of autocorrelations, of partials, and within the line spectrum, will exhibit a possible model.

ii. The graph of the data set is a visual assistance to identify the behaviour of the pattern. The autocorrelations and the line spectrum are the summary of the pattern existing in the data. They can reveal a great deal about the data and their characteristics.

iii. The model can be used in the present case to express the dependence between X_t and X_{t-1} in the pair (X_t, X_{t-1}), and to thus relate X_t with X_{t-1}, X_{t-1} with X_{t-2} and so on.. The plot of X_t and X_{t-1} for t = 2, 3, …, 296 is depicted in Figure 4. It can be examined that the points are scattered around a straight line. This straight line trend also depicts that X_t does depend on X_{t-1} (X_{t-1} on X_{t-2}, and X_{t-2} on X_{t-3}, and so on). We can write the model relating X_t and X_{t-1} as given in Eq. 4

$$X_t = \phi X_{t-1} + a_t$$

The above model expresses the dependence of the variable on itself at different times for model under consideration a_t at different t are independent, that a_t is independent of a_{t-1}, a_{t-2}, etc. so that just like ε_t the distribution of a_t is assumed to be normal,

$$a_t \sim NID(0, \sigma^2_a) \qquad (5)$$

It has been noted that estimated model is completely specified only when σ^2_a is given in addition to ϕ_1. a_t is assumed that it does not depend on the X_{t-2}, X_{t-3}.

The value of X_t may increase or decrease without bound, because a_t have fixed finite variance and can not continually increase in magnitude to keep X_t within bound as depicted from Figure 5 that explains the residual analysis specified for this model and also confirms that this model is adequate.

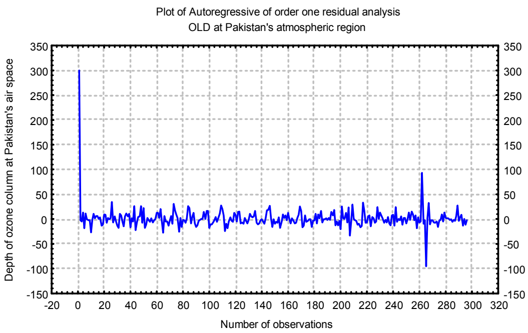

Figure 5. Plot of analysis the residuals (errors) for Autoregressive model of order one after fitting to the OLD data for atmospheric of Pakistan.

Such a situation tells that if ϕ_1 >1 or ϕ_1 < -1, then the series will be non-stationary or unstable time series. For a stationary stable time series, X_t remains bounded in the sense it has finite variance. we would need ϕ_1 <1. Figures 6 and 7 illustrate the estimated auto-correlation function and the auto-correlation function from lags 1 to 15 respectively. The autocorrelation plot represents an important characteristic of the linear stochastic model. Figure 6 depicts the estimated correlation between the i^{th} observation and the $(i + m)^{th}$ observation on y-axis vs the lag number on the x-axis. Similarly, Figure 7 shows autocorrelation plot for the residuals of the ozone layer depths.

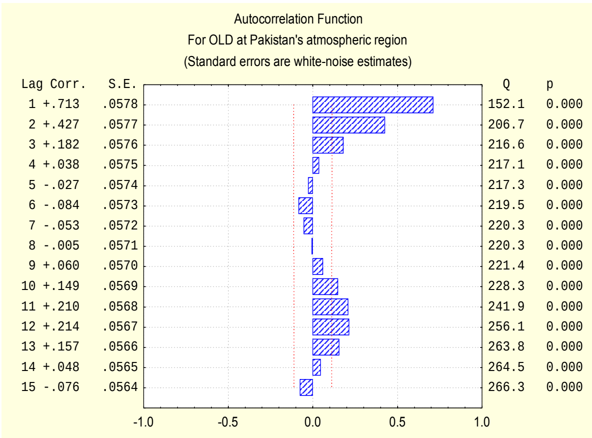

Figure 6. Plot of Autocorrelation function for OLD

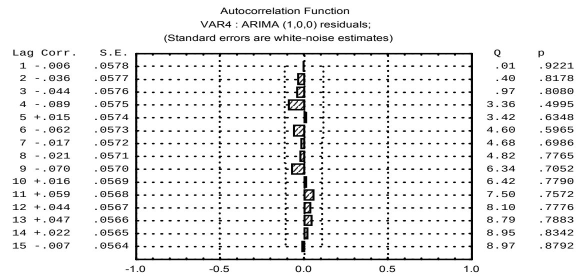

Figure 7. Autocorrelations for the residuals of OLD, n = 296.

The dashed lines in the above mentioned figures display an approximate 95 % confidence interval for an individual estimated autocorrelation. Thus, for independent observations, a particular confidence limits about 95 % of the time. In these cases, the autocorrelation plots strongly recommend the presence of serial correlation in the ozone data.

(e) Inspection of model adequacy: The constructed model can be inspected by dividing the entire data set into two parts. First section which is regarded as the major part of the data set is operated to estimate or to compute the parameters. The values of the parameters estimated known as the predicted values. These values are compared with the minor portion of the data set which are regarded as the observed values as depicted in Table 2. Observed values are plotted against predicted values as depicted in the (vide figure 8). These illustrations are verifying the results of estimates obtained from the estimating technique.

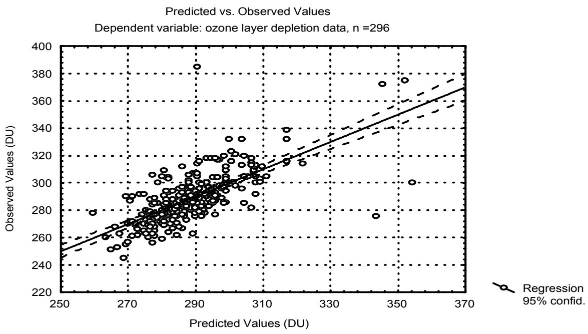

Figure 8. Plot of observed and predicted values verifying as a results of estimates obtained from the estimating techniques.

Table 2. Comparison of observed and predicted OLD values.

S.No.	Observed	Predicted
297	284.000	267.5927
298	268.000	267.1861
299	265.000	266.000
300	265.000	266.7801
301	262.000	265.9700
302	272.000	265.1600
303	272.000	264.7600
304	273.000	264.3600
305	289.000	263.9500
306	282.000	263.8756

(f) Computation of parametric values for data points (N = 296)

Following parameter tests were performed for the model verification (regression and stochastic)

a. Standard error test as computed from regression models and stochastic model.
b. MSE test (Mean Square test).
c. SSE test (Sum of the square test).
d. t-test (t- distribution test).
e. F-test (Fraser test).
f. p-value confidence interval (from +95% to -95% confidence level) for models

Table 3. Tests of statistical significance.

Model / Test	Simple Regression	Multiple Regression	Stochastic model AR-1
Estimate β_1	− 0.43	− 1.097	0.710±0.157
St.Error	9.35	1.48	0.032
T – statistics	− 1.09	− 0.018	127.084
Constant β_0	298.80	285.113	178.653±1.4
Estimate β_2	-	− 0.018	-
MSE	102.68	1772	78.29
SSE	102.68	3544	2975.18
$F_{0.95}$ – value	1.18	1.12	0.903
p – value	0.284	0.348	0.025
Corr.Coef.	Negative	Negative	Positive
R^2	35 %	27 %	51 %

Table 4. Tentative prediction for Pakistan using different modeling strategies.

Model	Simple Regression	Multiple Regression	Stochastic model AR (1)
Prediction for 2001	294.93 DU	283.42 DU	273.66 DU
MAFE or Forecast Accuracy	9.13	13.5	4.5 %

3. Conclusion

Using the prediction equation obtained from the least squares, we estimate the mean value of the depth or to predict a particular value of the depth of ozone column for different values of time that fall outside the range of values of time contained in the sample may lead to an error of estimation and prediction that are much larger than expected. Albeit the least squares model may provide a very good fit to data but provide an inadequate manifestation of the true model for values of time outside this range.

The autoregressive models are clearer and easier to handle than the moving average from the OLD data set which we have analyzed. In the case of finding the appropriate model for OLD, we have looked into the major parametric values of the models. The adequate model is justified for having insignificantly small residual autocorrelations. It can be seen that the AR (1) is suitable for making prediction and finding forecasts for Pakistan's atmospheric region. We have found and confirmed that the underlying processes of the phenomenon of OLD are stochastic. For the modeling of such processes we have attempted various autoregressive models from AR(1) – AR(8) and found that AR(1) model is the most adequate [13].

Acknowledgements

The author would like to thank the Dean Faculty of Science for carrying out this work. I also pay my gratitude to Director and Staff of the Geophysical Centre, Quetta for providing me the data recorded by the Dobson Spectrophotometer under the auspices of World Meteorological Organization (WMO).

References

[1] R. Garcia, Physics World **17** (1994) 49.

[2] M.A.K. Yousuf Zai and J. Quamar, Indian J. Phys. **75B**, No. 4 (2001) 307.

[3] NASA Facts, On Line, NASA Goddard Space Flight Centre, USA (2003).

[4] L.G. Richard and R. B. May, Environ. Sci. Technol. **4** (1988) 22.

[5] L. A. Alvin, Env. Sci. Technol. **3** (1993) 27.

[6] L .Mátyás (ed.), Generalised Method of Moments Estimation, UK, Cambridge, University Press (1999).

[7] H. Allen and R.Murphy, Probability, Statistics and Decision Making in Atmospheric Sciences, London: West View Press (1987).

[8] K.E. Trenber and J. G. Olson, Greenhouse-Gas-Induced Climate Change: A Critical Appraisal of Simulations and Observations, Vol. 31, Amsterdam : Elsevier (1991) pp. 249-259.

[9] M. R. K Ansari and J. Quamar in Proceedings of Fifth International Conference of Science, Technology & Development in the New Millennium held at University of Karachi, Pakistan, April 24 – 27 (2000) pp. 205-223.

[10] J. W. Tukey, Exploratory data analysis, Reading, Mass., Addison- Wesley (1977).

[11] S. M. Pandit and S.M. Wu, Time Series and System Analysis with Applications, New York: J. Wiley (1985).

[12] R. Bartoszynski and M. N. Niewiadomska-Bugaj, Probability and Statistical Inference, Wiley-Inter Science, New York (1996).

[13] M.A.K Yousuf Zai, Ph.D thesis submitted at the Institute of Space and Planetary Astrophysics, University of Karachi (2003).